PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2008 - 74

The network structure of mutual support links: Evidence from rural Tanzania

Margherita Comola

JEL Codes: C52, D85, O12
Keywords: Mutual support, network formation, structural estimation, indirect contacts

PARIS-JOURDAN SCIENCES ECONOMIQUES
LABORATOIRE D'ECONOMIE APPLIQUÉE - INRA

48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES – ÉCOLE NORMALE SUPÉRIEURE

The Network Structure of Mutual Support Links: Evidence from Rural Tanzania *

Margherita Comola †

June 2010

[Previous versions have circulated with the title:
"The Network Structure of Informal Arrangements: Evidence from Rural Tanzania"]

Abstract

This paper takes a network perspective to investigate how rural households in developing countries form the links through which they provide and get economic support. I test the hypothesis that indirect contacts (*e.g.* friends of friends) matter for link formation. An estimation procedure of a network formation model *à la* Jackson and Wolinsky (1996) is proposed and applied to data on a single village in Tanzania. Results show that when agents evaluate the net advantage of forming a link they also consider the wealth and the position of indirect contacts. The network externalities from indirect contacts are negative, which suggests a mechanism of competition over scarce resources. This paper proposes the first structural estimation of an endogenous network formation model, and also contributes to the development literature by overcoming the dyadic regression approach and providing evidence that village-level network structure has an explanatory value disregarded by all previous studies.

Keywords: Mutual Support; Network Formation; Structural Estimation; Indirect Contacts

JEL codes: C52; D85; O12

*I thank Joachim De Weerdt for making the data available, and Antonio Cabrales, Marcel Fafchamps, Yann Bramoullé and Marco van der Leij for their excellent comments.

†Paris School of Economics. Email: `comola@pse.ens.fr`

1 Introduction

As social capital theorists have stated for some time (Coleman, 1988; Wilson, 1987) and economists have realized more recently, a person's network of contacts constitutes a crucial resource. This is especially the case for rural communities in developing countries where all outcomes, from weddings to money lending, are determined by informal, multi-purpose interactions (Fafchamps and Lund, 2003; Hoddinott, Dercon and Krishnan, 2005). Whenever economic and financial institutions are weak, people in need are forced to rely on family, friends and associates; therefore social links assume an economic value. The most common example are risk-sharing arrangements: when households have no access to credit, either because there are no institutions that provide it or because households cannot meet the collateral required to enter a formal transaction, private transfers such as loans or gifts are used to stabilize consumption in face of idiosyncratic shocks as health-related expenses, funerals and court trials. However several other forms of network-based mutual support such as work parties, sharecropping and oxen sharing have been documented (Hoddinott, Dercon and Krishnan, 2005; Krishnan and Sciubba, 2009). In most circumstances people do not form links specifically for the economic outcome, but the economic arrangements originate from preexisting bilateral relationships (Fafchamps and Gubert, 2007a).[1] This paper focuses on the bilateral links at the basis of mutual support arrangements, rather than on their outcomes: I investigate how rural villagers form the links through which they provide and get economic support, and whether the structure of the community affects link formation. Specifically, I answer the question of whether agents choose their partners on the basis of their individual characteristics only, or also for their position with respect to the other agents? That is, do indirect contacts (*e.g.* friends of friends) matter?

Villages are the most common economic and social structure throughout the developing world, accounting for two-thirds of Sub-Saharan African population. For their limited spatial and informational barriers and their complex social interactions, rural villages are an ideal setting to study endogenous network formation. I base my analysis on a village called Nyakatoke, in the Buboka Rural District of Tanzania, west of Lake Victoria. In the Nyakatoke Households Survey all adult individuals were asked *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?"*; I use this piece of information to trace the architecture of the mutual support network, and then propose a structural estimation procedure to test whether indirect contacts matter.

[1] The same perspective is adopted in the model by Bramoullé and Kranton (2007), where individuals first form the network of bilateral relations and then use these relations to share income after income shocks are realized.

From the seminal research of Jackson and Wolinsky (1996) and onwards, economic network theory claims that the entire community graph is relevant for the formation and severance of links. However, none of the empirical studies on the determinants of link formation (De Weerdt, 2004; Dekker, 2004; Conley and Udry, 2005; Arcand and Fafchamps, 2008) acknowledges the importance of the community structure. The aim of my paper is to fill this gap. Borrowing from Jackson and Wolinsky (1996), in my framework links are formed at the discretion of self-interested agents whose utility is given by the overall network structure. Specifically, link formation is determined by the wealth and the social characteristics of the potential partner, and I allow externalities from indirect contacts in the dimension of wealth. I propose an estimation procedure and derive testable predictions on wealth externalities that I bring to data. Not surprisingly, I find that Nyakatoke villagers are more willing to link with wealthy partners. More interestingly, I also find that indirect contacts do matter, that is, that two potential partners with identical characteristics but with different social connections are not worth the same. Data show that wealth externalities from two-steps-away contacts are negative: agents prefer wealthy partners who have few/poor additional contacts, which suggests a competition mechanism. As network theory acknowledges from its very first steps (Jackson and Wolinsky, 1996), link formation may generate positive or negative externalities: indirect partners are beneficial if they broaden social interactions, but detrimental if there is competition for scarce social and/or economic resources. Mutual support arrangements are an intriguing example that combines positive and negative indirect externalities, and my results suggest that among Nyakatoke villagers the competition component prevails.

This paper contributes to both network and development economics. First, it proposes an innovative procedure to estimate endogenous network formation models, showing how partial observability models (Poirier, 1980) can be used to estimate pairwise stability conditions, and ultimately bilateral link formation. This approach allows me to overcome the dyadic model, and to separately model the willingnesses of the two partners to engage in a link. Second, this paper highlights the importance of indirect contacts to understand link formation. The only previous empirical study which explicitly recognizes a role for network architecture is the one by Krishnan and Sciubba (2009). However, while Krishnan and Sciubba test their theoretical prediction against data in a reduced form framework, I propose a structural analysis and recover the parameters of the individual utility function. From a policy point of view, the major lesson of this paper is that when agents choose their mutual support links they also look at potential partners' social connections. As Dasgupta (2003) points out, informal networks have effects that spill over to all areas of economic activity, so that understanding the unwritten rules driving informal ties is necessary to design social protection policies at the micro level.

The paper is organized as follows. Section 2 contains a review of the relevant literature. In Section 3 and 4 the data description and the theoretical framework are respectively presented. Section 5 introduces the estimation procedure, and Section 6 discusses the specifications and the results. Section 7 concludes summarizing the main findings. Tables and figures are presented in the Appendix at the end of the paper.

2 Literature Review

The economic literature on mutual support in developing countries is mainly empirical and focused on risk-sharing arrangements.[2] As Mace (1991) has pointed out, the optimal insurance scheme would be full income pooling at the village level, but this is not realized because of imperfect commitment and information asymmetry among villagers (Ligon, 1998; Coate and Ravallion, 1993; Ligon, Thomas and Worrall, 2000). Fafchamps and Lund (2003) investigate how households deal with shocks in rural Philippines concluding that mutual insurance takes place in small groups of friends and relatives. De Weerdt (2004) finds that the main variables predicting mutual support arrangements in Tanzania are kinship, distance, religion and common friends. Grimard (1997) and Fafchamps (2003) explicitly focus on the impact of ethnicity. Along similar lines we find the descriptive assessments by Hoddinott, Dercon and Krishnan (2005), Rosenzweig (1988), Conley and Udry (2005), Dekker (2004) and Goldstein, De Janvry and Sadoulet (2002). The empirical literature on link formation has mostly used the dyadic regression model, taking the pair of agents (*dyad*) as unit of analysis and restrict the covariates to relational variables, while the procedure I propose allows me to incorporate individual-specific regressors.

None of the existing contributions explicitly recognize a role for network structure, with the only exception of Krishnan and Sciubba (2009), who analyze labor exchange arrangements for harvesting and weeding in 15 Ethiopian villages. They modify the coauthor model by Jackson and Wolinsky (1996) allowing for heterogeneity among agents: externalities from indirect contacts are negative because a higher number of partners dilute the effort a farmer can exert in each partnership. They identify the common features shared by any stable network and then check whether the arrangements observed in data are compatible with the model's predictions. My paper departs from a similar setting but, while Krishnan and Sciubba test the theoretical features of stable equilibria, I recover the underlining parameters of the network formation model.

[2]Among the theoretical contributions Genicot and Ray (2003) study the effect of allowing subgroup deviation in risk-sharing arrangements and Bloch, Genicot and Ray (2008) characterize the properties of stable insurance schemes for exogenously given network structures.

My paper builds on the endogenous network formation literature that has flourished in the last few years (Jackson, 2004; Jackson, 2006). Models of strategic network formation have originated primarily from two sources: the random graph literature by physicists (Vega-Redondo, 2007; Guimera' *et al.*, 2003; Boguna' *et al.*, 2004) and the economic literature inspired by game theory. Some economists have approached network formation from a non-cooperative point of view (Bala and Goyal, 2000; Galeotti, Goyal and Kamphorst, 2006). However, the majority of them have focused on stable networks, where links are formed at the discretion of self-interested agents whose utility is given by the overall network structure. Previous studies on applied networks have assessed how the exogenous position of players in the network impacts the social outcome (Calvó-Armengol, Patacchini and Zenou, 2009; Conley and Topa, 2002; Conley and Udry, 2005), while to the best of my knowledge my paper is the first attempt to structurally estimate endogenous network formation.

3 Nyakatoke Household Survey

The data I use come from the Nyakatoke Household Survey. Nyakatoke is a small village of the Haya tribe in the Buboka rural district of Tanzania. The community is composed of 600 inhabitants, 307 of which are adults, for a total of 119 households. Inhabitants have been interviewed in five regular intervals from February to December 2000. First all household heads, and a few days later, all adults were interviewed; this has produced a rich dataset containing information on households' demographics, wealth, income sources and income shocks, transfers and relationships of mutual support. Even if some piece of information was collected at the individual level, in this paper the $119 * 118 = 14042$ household dyads are taken as units of analysis.

Mutual support relationships are crucial for Nyakatoke households, as they self-report transfers (*i.e.* zero-interest loans and gifts) to be the main strategy to cope with idiosyncratic shocks like sickness, death, crime and court cases, and ceremonies (Dercon and De Weerdt, 2006).[3] During the first survey round all adult households' members were asked *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?"*; this is the piece of information I use to define whether a mutual support link exists and to trace the

[3] In the fifth survey round (December 2000) all adults were asked to enumerate the two worst shocks experienced in the past 10 years. As Dercon and De Weerdt (2006) report, the respondents listed a total of 296 shocks: illness is the most frequently mentioned shock, followed by death/funeral, ceremony, and crime and court cases. Community-correlated events (such as bad agricultural prices or weather shocks) are mentioned only 12 times.

village architecture. Self-declared link data are open to different interpretations so that the definition of link necessarily involves some arbitrariness (Comola and Fafchamps, 2009). The answers to the question above describe the bilateral relationships at the basis of mutual support arrangements, rather than the actual economic outcomes (*i.e.* loans and gifts, labour sharing) which may or may not take place depending on the situation and the specific needs and income shocks of the partners. In this interpretation, I assume that declared links are symmetric (*i.e.* if i is a friend of j then j must be a friend of i) and the link formation process is bilateral (*i.e.* both parties involved need to agree for a link to be formed, even if realized transfers do not necessary flow in both directions). This poses however a problem for the dyads where i mentions j but j does not mention i: non-reciprocated declarations must be imputed to some sort of measurement errors, and it would be equally legitimate to assume under-reporting by j or over-reporting by i. For the way the question is formulated (somebody they can rely on *and/or* that can rely on them) and to facilitate the estimates convergence, I choose under-reporting, and as in De Weerdt (2004) every time a respondent mentions another one I draw a link between the two households they belong to. With this procedure 980 links within the 14042 households dyads are identified.[4] The resulting network is dense, with a mean geodesic distance (*i.e.* number of links in the shortest path between two partners) of 2.5 steps and a maximum geodesic distance of 5 steps. No household is isolated, and the number of households' reported links ranges from 1 to 32. The network exhibits all the empirical regularities of large social networks (small world properties) that have been described in the literature.[5] For a graphical representation of the Nyakatoke network see Appendix, Figure 1.A.

Nyakatoke's village area is small, with an average distance between households of 523

[4]The estimations in Section 6 have been repeated under the assumption of over-reporting (a link exists only if both sides declare it). Results are consistent in sign and magnitude but not always in significance, which is unsurprising given the exiguous number of links (280 instead of 980). This seems to suggest that the few reciprocated links are simply the better quality ones, that is, the ones where both sides are highly committed. Also, the very small number of positive outcomes exacerbates the convergence difficulties partial observability models are well known for.

[5]These typical features are the following: a) the number of nodes is very large as compared to the average number of links; b) the network has an unique component or a main component covering a large share of the population; c) there are more nodes with a really low or a really high number of connections than in a network where links are formed uniformly at random; d) the number of links of connected partners tends to be positively correlated; e) the average distance between nodes is small and the maximum distance (diameter) is in the order of $ln(n)$, (which is precisely our case since $ln(119) = 4.77$); f) the clustering coefficient, which measures the tendency of linked nodes to have common neighbors, is larger in social networks with respect to the case where links are generated by an independent random process (for Nyakatoke the clustering coefficient is 0.23, where in an analogous randomly generated network it would be 0.03). For further details on small world properties see Jackson and Rogers (2007).

meters and a maximum distance of 1738 meters. The village is relatively poor, with a consumption for adult equivalent unit of less than 2 US $ a week and an average food share in consumption of 77% (Dercon and De Weerdt, 2006). Households in Nyakatoke get most of their income from agricultural activities, especially the cultivation of coffee and banana; other sources of income are rare and off-farming activities are mostly considered supplementary to farming (Mitti and Rweyemamu, 2001). The village has no primary school (the closest one is at 2 km) and overall educational level is low, with 26 households out of 119 where no member has completed primary education. All households nowadays are either Muslim, Lutheran or Catholic, and the three main formal religious associations play an active role in the social and economic life of the village. Finally, In Nyakatoke there are 26 different clans, with a variable number of households from 1 to 23.

A detailed variable description for the empirical analysis will come with Section 6, and households attributes are described in the Appendix, table 1.A. For additional information on Nyakatoke I remand to Mitti and Rweyemamu (2001) and De Weerdt (2002).

4 The Theory

Let $N = \{1, ..., n\}$ be a set of players connected in some network relationship, where links are the consequence of agreement between parts. Let the network be denoted by the adjacency matrix $g = [g_{ij}]$: for any pair of players $\{ij\}$ such that $i, j \in N$ and $i \neq j$, $g_{ij} = 1$ indicates that i and j are linked under the network g, and $g_{ij} = 0$ otherwise. The network matrix is symmetric, that is, $g_{ij} = g_{ji}$. By a standard abuse of notation, let g_{+ij} denote the network g with the link g_{ij}, that is, with $g_{ij} = 1$ and, analogously, g_{-ij} defines the network g without the link g_{ij}, that is, with $g_{ij} = 0$. $N(g)$ is the set of players forming the network g.

In order to identify which networks are likely to arise in various contexts, a notion of network stability is needed. Let us write $u_i(g)$ the utility that agent i derives from network g. The pairwise stability notion by Jackson and Wolinsky (1996) states that the formation of a link requires the consent of both parties involved, while severance can be done unilaterally; formally, a network g is pairwise stable if

$$\begin{aligned} (i) \quad & \forall g_{ij} = 1, u_i(g_{+ij}) \geq u_i(g_{-ij}) \;\&\; u_j(g_{+ij}) \geq u_j(g_{-ij}) \\ (ii) \quad & \forall g_{ij} = 0, u_i(g_{-ij}) < u_i(g_{+ij}) \Rightarrow u_j(g_{-ij}) > u_j(g_{+ij}) \end{aligned} \tag{1}$$

That is, a network is pairwise stable if links which are *ceteris paribus* profitable for both parties are actually formed, and each player does not benefit in severing any existing link.

Pairwise stability does not admit multiple or simultaneous deviations. As it frequently admits large sets of stable allocations, several refinements have been proposed (Jackson and Van Den Nouweland, 2005; Jackson and Wolinsky, 1996) however, for the purpose of this paper the pairwise stability concept will be adopted.

The connection model and the co-author model are two concrete network examples presented in Jackson and Wolinsky (1996). In the connection model, links represent social relationships, which provide benefits but also involve costs. Players incur a cost for every link they form; on the other hand, they get positive externalities from indirect contacts, which are for free. Benefits from indirect contacts deteriorate with distance: a friend is more valuable than a friend of a friend, and so on. In the co-author model, agents are interpreted as researchers who collaborate in common projects, and network externalities are negative: the amount of time a researcher can spend on any given project is inversely proportional to the number of projects he is involved in, so that indirect contacts are detrimental. Jackson and Wolinsky focus on simplified, symmetric versions of the general setting with homogeneous costs and benefits, and show that decentralized decisions do not necessarily lead to an allocation that maximizes collective utility.

I work with a setting of fully heterogeneous individuals instead: for any pair of agents $\{ij\}$ belonging to a community, let the m-dimensional vector Z_{ij} define their relative social characteristics. Let us define t_{ij} as the geodesic distance between i and j, that is, the number of links in the shortest path between then. Geodesic distance is symmetric ($t_{ij} = t_{ji}$) and is set to infinity if there is no path between i and j. The utility that agent i derives from network g is given by

$$u_i(g) = y_i + \sum_{j \in N(g)} a(t_{ij})\, y_j + \sum_{j: g_{ij}=1} \beta Z_{ij} \tag{2}$$

where y_i is i's wealth, $a(.)$ is the weight attached to the wealth of all other agents in the community as a function of their geodesic distance, and $\beta \in R^m$ are the weights attached to social characteristics of direct links only. In this simple linear parametrization, which generalizes the connection model, wealth of indirect contacts generates externalities, which can be positive or negative in sign. I then proceed to empirics, and propose an estimation procedure to test for the presence of those externalities.

5 Estimation Procedure

Using equation (2), for a network configuration g we can write

$$
\begin{aligned}
u_i(g_{+ij}) - u_i(g_{-ij}) &= \sum_{k \in N(g_{+ij})} a(t_{ik}) y_k - \sum_{k \in N(g_{-ij})} a(t_{ik}) y_k + \beta Z_{ij} && (3)\\
&= \sum_{\phi}^{n-1} \sum_{k} a(\phi) \left[y_{\substack{k \in N(g_{+ij}) \\ t_{ik}=\phi}} - y_{\substack{k \in N(g_{-ij}) \\ t_{ik}=\phi}} \right] + \beta Z_{ij} && (4)\\
&= a(1) y_j + a(2) \sum_k \left[y_{\substack{k \in N(g_{+ij}) \\ t_{ik}=2}} - y_{\substack{k \in N(g_{-ij}) \\ t_{ik}=2}} \right] \\
&\quad + a(3) \sum_k \left[y_{\substack{k \in N(g_{+ij}) \\ t_{ik}=3}} - y_{\substack{k \in N(g_{-ij}) \\ t_{ik}=3}} \right] + ... + \beta Z_{ij} && (5)
\end{aligned}
$$

Equation (5) shows how the overall net gain in terms of partners' discounted wealth can be decomposed by geodesic distance. Let us set $a(\phi) = 0$ for $\phi \geq 4$,[6] and call

$$
\begin{aligned}
\sum_k \left[y_{\substack{k \in N(g_{+ij}) \\ t_{ik}=2}} - y_{\substack{k \in N(g_{-ij}) \\ t_{ik}=2}} \right] &\equiv \textit{two steps gain}_{ij} \\
\sum_k \left[y_{\substack{k \in N(g_{+ij}) \\ t_{ik}=3}} - y_{\substack{k \in N(g_{-ij}) \\ t_{ik}=3}} \right] &\equiv \textit{three steps gain}_{ij}
\end{aligned}
$$

two steps gain$_{ij}$ (*three steps gain*$_{ij}$) expresses the net gain in terms of two (three) steps away agents' wealth that i gets if the link g_{ij} is formed and the rest of the network is kept constant. Those quantities can be positive or negative, and can be directly computed on data.[7]

To calculate *two steps gain* and *three steps gain* I need to define the reference network g first. Using the outcomes observed in data (*i.e.* constructing g_{+ij} and g_{+ij} for the observation $\{ij\}$ on g_{km} for all $\{km\} \neq \{ij\}$) is unsatisfactory for two reasons: first, it does not

[6] It needs to be so because in the Nyakatoke data the maximum geodesic distance is 5, which implies that the regressors accounting for gains from the 1^{st} to the 5^{th} step sum up to zero. Since the network is rather dense (with an average geodesic distance of 2.5), these readjustments are reabsorbed quickly: five-steps-away gains equal zero in 13816 out of 14042 observations. Thus gains up to the 4^{th} step are (almost perfectly) collinear, and I omit four-steps-away gains taking it as a reference category.

[7] Since the formation of one link changes the entire architecture, the patterns of those quantities are difficult to predict. For instance, if i sets a link with a j who used to be two steps away, the one-step-away gain of y_j is reflected in a two-steps-away loss, so that *two steps gain*$_{ij}$ is negative if no other two-steps-away gain (*i.e.* a third player who used to be further away and is now two-steps-away from i) counterbalances the loss of j.

address omitted variable endogeneity.[8] Second, we fall into the problem of inconsistency of simultaneous equations with binary variables pointed out by Maddala (1983).[9] In order to overcome both problems at once, I proceed in the following way:

1. I first run a dyadic probit regression to identify which relational characteristics predict link formation, in a way which is familiar to the risk-sharing literature (De Weerdt 2004, Fachamps and Gubert, 2007b).[10] I use it to construct a distribution of possible networks $\widehat{p}$, where the entry $\widehat{p}_{ij}$ is the fitted probability of the link $g_{ij} = 1$ obtained from the dyadic regression.

2. From the distribution $\widehat{p}$ I draw 20 networks, which I call $\widehat{g}^1, ..., \widehat{g}^{20}$.

3. For each of these generated networks $\widehat{g}^1, ..., \widehat{g}^{20}$ I compute my regressors of interest *two steps gain* and *three steps gain*, which is a matter of advanced matrix manipulation.[11]

4. The final values of *two steps gain* and *three steps gain* included in the regressions are the average across these 20 values, which hopefully minimizes the sampling variability.

The regressors obtained from this procedure are exogenous but correlated to the observed network and hence to the true network externalities we want to measure. The interpretation I have in mind is a setting of imperfect information where individuals share a common belief about which links are more likely to be formed, however the framework can be adapted to fit

[8] Omitted variable endogeneity is crucial for inference in network formation: if agents a and b are observationally equivalent but a has many links while b has none, it may be either because a grants access to many indirect contacts as I would like to believe, or for some unobserved a's characteristics.

[9] Essentially Maddala shows that, if we condition an individual binary outcome on the action undertaken by another individual in the sample, the likelihood is not valid because it does not sum up to one over all possible states (network configurations in this case). See also Tamer (2003).

[10] I run a probit regression over the full sample (14042 dyads), where the binary dependent variable equals one if a link is declared by any of the two households involved (Section 3). The regressors included are standard (Fachamps and Gubert, 2007b): sum and absolute difference of land and livestock wealth, distance, dummies for kinship, same religion, same clan, overlap in productive activities of the two households, and education (two dummies equal to one if both households have some member who completed primary education or both households have no member who completed primary education respectively). For a detailed definition of those variables see Subsection 6.1.

[11] For any of those generated networks, say $\widehat{g}^1$, I need to: a) generate the network matrices $\widehat{g}^1_{+ij}$ and $\widehat{g}^1_{-ij}$ for each dyad $\{ij\}$, which makes 14042*2 matrices that need to be created; b) generate the associated 14042*2 matrices of geodesic distances using an algorithm similar to the one proposed by Dijkstra (1959); c) combine those matrices with the vector of individuals' wealth: for each $\{ij\}$, the quantity *two steps gain*$_{ij}$ is computed as the sum of the wealth of all $k \in N(\widehat{g}^1_{+ij})$ such that $t_{ik} = 2$ minus the sum of the wealth of all $k \in N(\widehat{g}^1_{-ij})$ such that $t_{ik} = 2$. Similarly, *two steps gain*$_{ji}$ is computed as the sum of the wealth of all $k \in N(\widehat{g}^1_{+ij})$ such that $t_{jk} = 2$ minus the sum of the wealth of all $k \in N(\widehat{g}^1_{-ij})$ such that $t_{jk} = 2$.

different stories.[12]

Let us define a response variable w_{ij} which represents i's willingness to form the link g_{ij}

$$w_{ij} = \begin{cases} 1 & \text{if } u_i(g_{+ij}) - u_i(g_{-ij}) + \epsilon_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where ϵ_{ij} is a zero-mean residual. Analogously, w_{ji} represents j's willingness to form the link g_{ij}

$$w_{ji} = \begin{cases} 1 & \text{if } u_j(g_{+ij}) - u_j(g_{-ij}) + \epsilon_{ji} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Under pairwise stability (Equation 1), link formation reduces to

$$\begin{aligned} P(g_{ij} = 1) &= P\left(w_{ij} = 1 \;\&\; w_{ji} = 1\right) \\ P(g_{ij} = 0) &= 1 - P\left(w_{ij} = 1 \;\&\; w_{ji} = 1\right) \end{aligned} \tag{8}$$

This two-equation system is estimated as a bivariate probit with partial observability *à la* Poirier (1980) (see also Maddala, 1983; Abowd and Farber, 1982; Farber, 1983). In the standard bivariate probit, a joint distribution of the errors $[\epsilon_{ij}, \epsilon_{ji}] \sim$ bivariate normal $[0, 0, 1, 1, \rho]$ is assumed. Partial observability occurs when we can observe a positive outcome for one response variable only if the other response variable is also positive. In other words, we do not fully observe w_{ij}, w_{ji} but when $g_{ij} = 1$ we know that $w_{ij} = w_{ji} = 1$. *Viceversa*, when $g_{ij} = 0$ we cannot distinguish between the following three cases: $w_{ij} = 0$ & $w_{ji} = 1$; $w_{ij} = 1$ & $w_{ji} = 0$; $w_{ij} = w_{ji} = 0$. Provided that the system in properly identified, evaluating the associated likelihood function we can estimate the two equations and recover the determinants of w_{ij} and w_{ji} separately.

Therefore, using Equation (5) for each $\{ij\}$ I estimate the two equations

[12]For instance, we can imagine agents sophisticated enough to compute from $\widehat{p}$ the exact probability of all network configurations, which makes the construction of *two steps gain* and *three steps gain* computationally more demanding but avoids the multiple networks draws. On another line, we can assign agent-specific beliefs, which may be useful to investigate issues related to equilibrium refinement. Also, instead of using *two steps gain* and *three steps gain* as exogenous regressors, we could use them as instrumental variables in a traditional two-stage strategy (results with bootstrapped standard errors are similar to the ones in Subsection 6.2.).

$$w_{ij} = \begin{cases} 1 & \text{if } a(1)y_j + a(2)\ \textit{two steps gain}_{ij} + a(3)\ \textit{three steps gain}_{ij} + \beta Z_{ij} + \epsilon_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$w_{ji} = \begin{cases} 1 & \text{if } a(1)y_i + a(2)\ \textit{two steps gain}_{ji} + a(3)\ \textit{three steps gain}_{ji} + \beta Z_{ji} + \epsilon_{ji} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the estimated coefficients a and β are the same by construction across the two equations because of the data symmetry (*i.e.* both dyads $\{ij\}$ and $\{ji\}$ are included). This provides a straightforward test of my hypothesis: if the coefficients $a(2), a(3)$ turn out significant this suggests that when agents form arrangements they also consider the relative position and the wealth of indirect contacts.

One last issue to be solved is the correlation of residuals among observations. As the sampled individuals are the same across dyads, the dyadic observations are by construction non-independent, which invalidates the standard errors. I specify a clustered covariance matrix as the one in Fafchamps and Gubert (2007a), which corrects for arbitrary correlation across all observations involving either i or j: the only structure imposed on the covariance is that $E[\epsilon_{ij}, \epsilon_{ik}] \neq 0$, $E[\epsilon_{ij}, \epsilon_{kj}] \neq 0$, $E[\epsilon_{ij}, \epsilon_{jk}] \neq 0$ and $E[\epsilon_{ij}, \epsilon_{ki}] \neq 0$ for all k but that $E[\epsilon_{ij}, \epsilon_{km}] = 0$ otherwise. This formula also corrects for heteroskedasticity and for the double counting of the dyads $\{ij\}$ and $\{ji\}$.

6 Specifications and Results

6.1 Variable Definition

For each dyadic observation $\{ij\}$ we separately estimate two binary equations, whose dependent variables w_{ij} and w_{ji} are partially observed through g_{ij} (defined in Section 3). We report the estimated results for w_{ij}, that is, the willingness of i to form the link g_{ij}. Results for w_{ji} are identical because of the data symmetry, and therefore are not reported.

The willingness to form a link w_{ij} is assumed to depend on the wealth of j and of indirect contacts, and on the social characteristics of i and j. In particular:

$wealth_j$ is given by the total monetary value of j's land and livestock assets, while wealth gains from indirect contacts $\textit{two steps gain}_{ij}$ and $\textit{three steps gain}_{ij}$ are computed as explained in Section 5. The unit of measure are Tanzanian Shillings (1 unit=100000 *tzs*).[13] Most wealth is inherited through strict patrilineal rules and passed down the clan (customary land tenure laws used to prohibit the selling of land – see Mitti and Rweyamamu, 2001), which should reasonably rule out inverse causality between wealth and links. Still, part of the estimated effect in Section 6.2 may be seen as correlation rather than causation, and unfortunately no existing data would address the issue.

To proxy for the desire to link with similar households (homophyly),[14] I include the following relational dummies: *same clan* and *same religion* (which equal one if the two households belong to the same clan or profess the same religion respectively), *neighbours* (which equals one if the two households' houses are less than 300 meters apart), *kinship* (which equals one if a member of one household has a strict blood bond –child, parents, siblings– with a member of the other household) and *same education* (which equals one if both households are educated or both are non-educated, considering a household educated if at least one member has completed primary school).

To proxy for i's attitude and desire to link I include two religious dummies which equal one if household i is Lutheran or Catholic respectively (where Muslim is the omitted category), and the shares of households i's total labor devoted to cropping, casual labor and trade, which are the three most common income generating activities in Nyakatoke.[15] These latter variables are meant to control for the nature of i's income, as some activities guarantee a more stable income against non-idiosyncratic shocks.

In Nyakatoke mutual support links between wealthy and poor households are very frequent.[16] This suggests that wealthy households may have non-economic motivations to form

[13] Data on land were originally in acres, and were transformed in monetary equivalent with a conversion rate of 300000 *tzs* for 1 acre, which reflects average local prices in 2000. For international comparisons, the exchange rate in 2000 was 1 US dollar for 800 *tzs*.

[14] Social proximity between partners decreases enforcement and monitoring costs of links. Studying microfinance group-lending programs in the Andes Karlan (2007) concludes that more homogeneous groups have higher repayment rates due to their higher social capital and ease of monitoring.

[15] In the survey each adult mentions the productive activities he is involved in (coded in seven categories: casual labor, trade, crops, livestock rearing, assets, processing of agricultural products and other off-farm work). Since individuals do not mention the relative importance of each activity, all mentioned activities are assumed to contribute in the same measure. Therefore, for each household the share of total time devoted to cropping is calculated on the base of the number of active adult members and their activities.

[16] At first we could think that those links are always declared by the poorer partner, but this is not the case: when the link is declared by one side only, it is the wealthier partner in 42% of cases. Moreover, in the survey individuals could list as many contacts as they want and, since on average wealthy people also have more links (Table 5.A), they are likely to quote only the most important ones.

links, like altruism or symbolic attributes such as social esteem an political power.[17] To capture these effects, we include among the regressors the absolute value of the partners' wealth difference $difference\ wealth$ and a dummy $more\ wealth_i$ which equals one if household i is wealthier that j.

Descriptive statistics can be found in the Appendix, figure 2.A and tables 2.A to 3.A.

6.2 Results

Table 2 reports the results for the w_{ij} equation under three different specifications: column (1) refers to the benchmark case where only direct partners' wealth is taken into account for link formation. In columns (2) and (3) I include the net wealth gains from indirect contacts two and three steps away respectively. The corresponding marginal effects are reported in the Appendix, Table 4.A.

$wealth_j$ is positive and significant across all specifications: as expected, the wealthier a potential partner, the more desirable a link with him. $two\ steps\ gain_{ij}$, which represent the net wealth gain in term of two steps away partners if the link is formed, is significant and negative in sign. This negative externality seems to suggest a competition mechanism over scarce resources. On the other hand, $three\ steps\ gain_{ij}$ is not significant, suggesting that agents do not take into account the network configuration further than two steps away from them, perhaps because it would require too much information to do so. The absolute magnitude of direct and indirect effects is decreasing in the geodesic distance, as theory predicts. Subsection 6.3 elaborates on the interpretation of those results.

All other variables have the expected sign, and reconfirm what has been documented by previous studies on informal insurance arrangements (Fafchamps and Lund, 2003; De Weerdt, 2004; Dekker, 2004). Geographical proximity between two households significantly correlates with the existence of a link, and the same holds for kinship ties which is the variable with the highest impact. Clan belonging is marginally significant in specifications (2) and (3), with a positive sign. Education and i's attitude (religious belonging and income generating activities) do not seem significant. While i's religion is not significant, a common religion has a marginally significant impact on the willingness to form a link in specifications (2) and (3). $more\ wealth_i$ and $difference\ wealth$ are not significant, reconfirming that rich individuals are not reluctant to form links with the villagers who have less.

[17]Several economists have indeed interpreted gift transactions as exchange rather than altruism, and have recognized asymmetric exchanges as the base of village-level patronage relationships in developing countries (Breman, 1974; Scott, 1976; Fafchamps, 1992; Platteau, 1995).

Table 2: Results

dependent variable: w_{ij}			
	(1)	(2)	(3)
$wealth_j$	0.057**	0.067***	0.071***
	(0.027)	(0.016)	(0.018)
$two\ steps\ gains_{ij}$		-0.023***	-0.022***
		(0.003)	(0.003)
$three\ steps\ gains_{ij}$			0.004
			(0.004)
$neighbours$	0.472*	0.539***	0.549***
	(0.275)	(0.078)	(0.095)
$same\ clan$	0.186	0.202*	0.198*
	(0.141)	(0.106)	(0.110)
$Lutheran_i$	0.061	-0.058	-0.046
	(0.122)	(0.117)	(0.133)
$Catholic_i$	-0.013	-0.108	-0.103
	(0.111)	(0.105)	(0.115)
$same\ religion$	0.149	0.173**	0.172**
	(0.096)	(0.067)	(0.070)
$kinship$	1.450***	1.307***	1.311***
	(0.428)	(0.132)	(0.137)
$same\ education$	-0.007	0.004	0.009
	(0.040)	(0.058)	(0.059)
$labour\ share\ cropping_i$	-0.089	-0.200	-0.199
	(0.122)	(0.138)	(0.150)
$labour\ share\ casual_i$	-0.172	-0.037	-0.084
	(0.203)	(0.214)	(0.267)
$labour\ share\ trade_i$	-0.299	-0.152	-0.177
	(0.327)	(0.219)	(0.266)
$difference\ wealth$	-0.026	0.004	0.003
	(0.023)	(0.009)	(0.009)
$more\ wealth_i$	3.211	-0.886	-0.549
	(2.532)	(0.644)	(1.171)
$constant$	-1.751**	-0.510	-0.809
	(0.681)	(0.691)	(1.247)
$arctan(rho)$	-1.004	0.534	0.749
	(1.334)	(0.616)	(1.269)
N	14042	14042	14042

Note: dyadic-corrected standard errors in parenthesis.

The list of regressors in Table 2 is conservative because the dataset is rather small, and because partial observability models are very demanding from the estimation perspective, and well known for their convergence difficulties. Switching to a stepping algorithm for non-concave regions of the likelihood function alleviates part of the problem, but with certain large or non-differentiated sets of regressors convergence may not be achieved. However, the example is meant to be illustrative of the methodology proposed, and all the major findings are robust to alternative specifications, including the elimination of wealth outliers.

6.3 Discussion and Extensions

As network theory acknowledges from its very first steps (Jackson and Wolinsky, 1996), link formation may generate positive or negative externalities: indirect contacts are beneficial if they broaden social interactions, but they are detrimental if there is competition for scarce (social and/or economic) resources. Mutual support arrangements are an intriguing example combining positive and negative indirect externalities, and without doubt is the most important example in the context of rural developing countries. The estimates of $a(2)$ in Table 2 suggest that the villagers of Nyakatoke prefer rich partners, but with fewer *and/or* poorer contacts. Since the distribution of wealth is relatively egalitarian, with a Gini coefficient of 0.43 (Figure 2.A), it comes natural to interpret this result as a condition on the number (rather than on the wealth) of indirect contacts: a desiderable partner is wealthy, and has few additional friends. This interpretation, which stresses the scarcity of economic resources, is in line with the fact that most households in Nyakatoke live at the subsistence level (with an average food share on consumption of 77%), and that most benefits that can be extracted from partners are exclusive by their own nature, as anecdotal evidence reports.[18] However, the result can also be seen as a condition on the wealth of indirect contacts: holding the number of indirect contacts fixed while increasing their wealth, *two steps gains* increases and thus the willigness to form a link decreases. This cannot be seen as a pure economic motivation, as rich indirect contacts cannot be detrimental if we interpret mutual support links as access to insurance. However it may be intepreted as a motivation related to social norms, as patron-client relations and social esteem are strong determinants of transfers in agrarian societies (Platteau, 1995 and 1996).

It should be pointed out that agents do not need to be aware of the full network architecture in order for indirect contacts to matter as previously shown. For instance, it may be

[18] Mitti and Rweyamamu (2001) report that *"free gift of land for cultivation of seasonal crops among friends and neighbours is quite common. A little of the produce may be given to the landowner as a gesture of appreciation for the use of his land"* and *"some villagers (especially women) perform vigodi* [a term used in the locality to mean casual labor] *or working on other households farms for cash"*.

simply the case that partners with few additional contacts tend to be of a greater help in case of need, and that survey respondents internalize this. What these results show, which is innovative, is that the networks architecture has an explanatory value that we disregard if we only focus on a reduced-form estimation of direct characteristics. Previous papers on link formation have taken the dyad as unit of analysis to focus on relational variables only (Fafchamps and Gubert, 2007a; De Weerdt, 2004; Dekker, 2004). This paper is the first attempt to overcome the dyadic approach and disentangle the link formation process into two separate components modeling the decisions of the agents involved. This allows me to include individual-specific regressors and, among them, the proxies for the indirect gains *two steps gain* and *three steps gain*, which should be interpreted as revealed preferences over partners' network position (*"ceteris paribus, I prefer friends who have few/poor other friends"*).

As a final note, in the current specification I only allow externalities from indirect contacts in the dimension of wealth. However, my method can be extended to estimate a generalized payoff function of the type

$$u_i(g) = y_i + \sum_{j \in N(g)} [a(t_{ij})\, y_j + \beta(t_{ij}) Z_{ij}] \tag{9}$$

which would show which social characteristics have positive externalities, and whether the negative externalities for wealth are outweighed by positive social externalities.

7 Conclusions

This paper performs a structural estimation of a model of endogenous network formation. I approach mutual support arrangements from a network perspective, investigating how links are formed and whether indirect contacts generate externalities: I test the hypothesis that agents also consider the wealth of indirect contacts when they decide to form a link. A protocol to estimate network formation models *à la* Jackson and Wolinsky (1996) is proposed and applied to data on the Tanzanian village of Nyakatoke. Results show that indirect contacts matter, that is, that agents forming links take into account the structure of the community network. Network externalities from indirect contacts' wealth are shown to be negative, which suggests a mechanism of competition over scarce economic or social resources. This paper contributes to both applied network economics and the literature on informal arrangements in that it proposes an innovative procedure to estimate endogenous network formation models, and also provides evidence that indirect contacts have an explanatory value disregarded by all previous studies, which are focused on direct relations only.

References

Abowd, J.M and H.S. Farber (1982): "Job Queues and the Union Status of Workers" *Industrial and Labor Relations Review*, 35, 3, 354-367.

Arcand, J.-L. and M. Fafchamps (2008): "Matching in Community-Based Organizations", (mimeograph).

Bala, V. and S. Goyal (2000): "A Noncooperative Model of Network Formation", *Econometrica*, 68, 5, 181-1229.

Bloch, F., G. Genicot and D. Ray (2008): "Informal Insurance in Social Networks", *Journal of Economic Theory*, 143(1), 36-58.

Bogugna', M., R. Pastor-Satorras , A. Diaz-Guilera and A. Arenas (2004): "Models of Social Networks based on Social Distance Attachment", *Physical Review* 70.

Bramoullé, Y. and R. Kranton (2007): "Risk Sharing Networks", *Journal of Economic Behavior & Organization*, 64(3-4), 275-294.

Breman, J. (1974): *"Patronage and Exploitation-Changing Agrarian Relations in South Gujarat"*, University of California Press, Berkeley, CA.

Calvó-Armengol, A., E. Patacchini and Y. Zenou (2009): "Peer Effects and Social Networks in Education", *Review of Economic Studies*, 76(4), 1239-1267.

Coate, S. and M. Ravallion (1993): "Reciprocity without Commitment: Characterization and Performance of Informal Insurance Arrangements", *Journal of Development Economics*, 40, 1-24.

Comola, M and M. Fafchamps (2009): "Testing Unilateral versus Bilateral Link Formation", Paris School of Economics WP 2009-30.

Conley, T. and G. Topa (2002): "Socio-Economic Distance and Spatial Patterns in Unemployment", *Journal of Applied Econometrics*, 17, 303-327.

Conley, T. and C. Udry (2005): "Learning about a new technology: pineapple in Ghana", Federal Reserve Bank of San Francisco Proceedings.

Coleman, J.S. (1988): "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, 94, S95-S120.

Dasgupta, P. (2003): "Social Capital and Economic Performance: Analytics" published in *Foundations of Social Capital*, E. Ostrom and T.K. Ahn, eds (Cheltenham UK: Edward Elgar).

Dekker, M. (2004): "Risk sharing in rural Zimbabwe: An empirical Analysis of Endogenous Network Formation", mimeograph.

Dercon, S. and J. De Weerdt (2006): "Risk Sharing Networks and Insurance against Illness", *Journal of Development Economics*, 81, 2, 337-356.

De Weerdt, J. (2002): "Community Organizations in Rural Tanzania: A Case Study of the Community of Nyakatoke, Buboka Rural District", Economic Development Initiatives mimeograph.

De Weerdt, J. (2004): "Risk Sharing and Endogenous Network Formation", *Insurance Against Poverty*, Oxford University Press, chapter 10.

Dijkstra, E.D (1959): "A note on two problems in connexion with graphs", *Numerische Mathematik* 1, 269-271.

Fafchamps, M. (1992): "Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy" *Economic Development and Cultural Change*, 41 (1), 147-174.

Fafchamps, M. (2003): "Ethnicity and Networks in African Trade", *Contributions to Economic Analysis & Policy*: Vol. 2: No. 1, Article 14.

Fafchamps, M. and F. Gubert (2007a): "Risk-Sharing and Network Formation" *American Economic Review Papers and Proceedings*, 97, 75-79.

Fafchamps, M. and F. Gubert (2007b): "The Formation of Risk Sharing Networks" *Journal of Development Economics*, 83(2):326.50.

Fafchamps, M. and S. Lund (2003): "Risk Sharing Networks in Rural Philippines" *Journal of Development Economics*, 71,2, 261-287 (27).

Farber, H.S. (1983): "The Determination of Union Status of Workers", *Econometrica*, 51 (5), 1417-1438.

Galeotti, A., S. Goyal and J. Kamphorst (2006): "Network Formation with Heterogeneous Players", *Games and Economic Behavior*, 54(2), 353-372.

Genicot, G. and D. Ray (2003): "Group Formation in Risk Sharing Arrangements", *Review of Economic Studies* 70, 87-113.

Goldstein, M., De Janvry, A. and E. Sadoulet (2002): "Is a Friend in Need a Friend Indeed? Inclusion and Exclusion in Mutual Insurance Networks in Southern Ghana", STICERD Working Paper DEDPS 32.

Grimard, F. (1997): "Households Consumption Smoothing through Ethnic Ties: Evidence form Core D'Ivoire" *Journal of Development Economics*, 53, 391-422.

Guimera', R., L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas (2003): "Self-similar Community Structure in a Network of Human Interactions", *Physical Review* 68.

Hoddinott, J., S. Dercon and P. Krishnan (2005): "Networks and Informal Mutual Support in 15 Ethiopian Villages", Mimeograph.

Jackson, M. (2004): "A Survey of Models of Network Formation: Stability and Efficiency", in *Group Formation in Economics; Networks, Clubs and Coalitions*, Cambridge University Press.

Jackson, M. (2006): "The Economics of Social Networks", in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Cambridge University Press.

Jackson, M. and B. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random are Social Networks?" *American Economic Review*, 97(3), 890-915.

Jackson, M. and A. Van Den Nouweland (2005): "Strongly Stable Networks", *Games and Economic Behavior*, 51 (2), 420-444.

Jackson, M and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, 71,44-74.

Karlan, D. S. (2007): "Social Connections and Group Banking", *Economic Journal*, 117(517), F52-F84.

Krishnan, P. and E. Sciubba (2009): "Links and Architecture in Village Networks", *Economic Journal*, 119(537), 917-949.

Ligon, E. (1998): "Risk Sharing and Information in Village Economies", *Review of Economic Studies* 65, 847-864.

Ligon, E., J.P Thomas and T. Worrall (2000): "Mutual Insurance, Individual Savings and Limited Commitment", *Review of Economic Dynamics*, 3, 216-246.

Mace, B.J. (1991): "Full Insurance in the Presence of Aggregate Uncertainty", *Journal of Political Economy*, 99(5), 928-956.

Maddala, G.S. (1983): *"Limited Dependent and Qualitative Variables in Econometrics"*, Cambridge University Press.

Mitti R. and T. Rweyemamu (2001): "A Socio-Economic Profile of Nyakatoke", Mimeograph.

Poirer, D.J. (1980): "Partial Observability in Bivariate Probit Models" *Journal of Econometrics*, 12, 209-217.

Platteau, J.P. (1995): "An Indian Model of Aristocratic Patronage", *Oxford Economic Papers* 47 (4), 636-662.

Platteau, J.-P. (1996): "Traditional Sharing Norms as an Obstacle to Economic Growth in Tribal Societies", Technical report, CRED, FacultÈs Universitaires Note-Dame de la Paix, Namur, Belgium. Cahiers No. 173.

Ravallion, M. and Chaudhuri, S. (1997): "Risk and Insurance in Village India: Comment", *Econometrica*, 65(1), 171-184.

Rosenzweig, M.R. (1988): "Risk, Implicit Contract and the Family in Rural Areas of Low-income Countries" *Economic Journal*, 98, 393, 1148-1170.

Scott, J. (1976): *"The Moral Economy of the Peasant"*, Yale University Press, New Haven, CT.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *Review of Economic Studies*, 70, 147-167.

Udry, C. (1994): "Risk and Insurance in a Rural Credit Market: an Empirical Investigation in Northern Nigeria", *Review of Economic Studies*, 61, 495-526.

Vega-Redondo, F. (2007): *"Complex Social Networks"*, Econometric Society Monographs, Cambridge University Press.

Wilson, W.J. (1987): *"The Truly Disadvantaged: the Inner City, the Underclass, and Public Policy"*, University of Chicago Press.

Appendix

Figure 1.A: Links among Nyakatoke Households

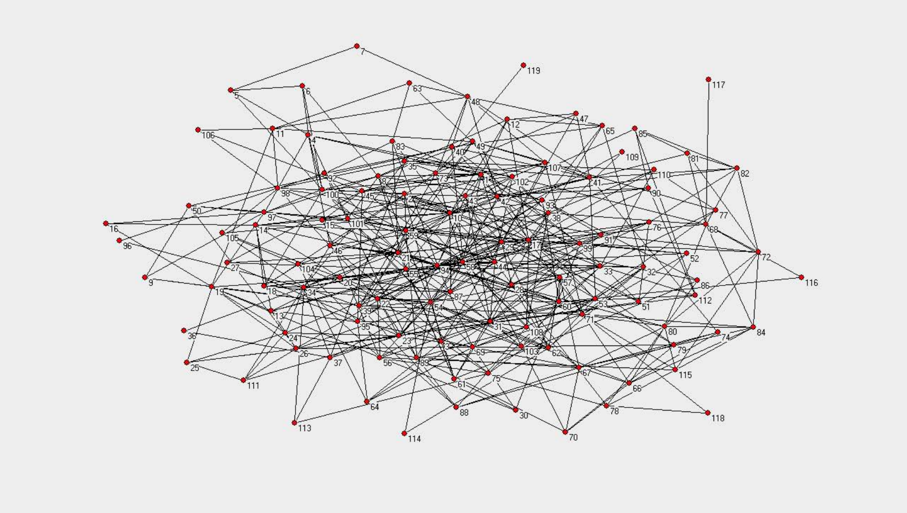

Figure 2.A: Wealth Quantiles

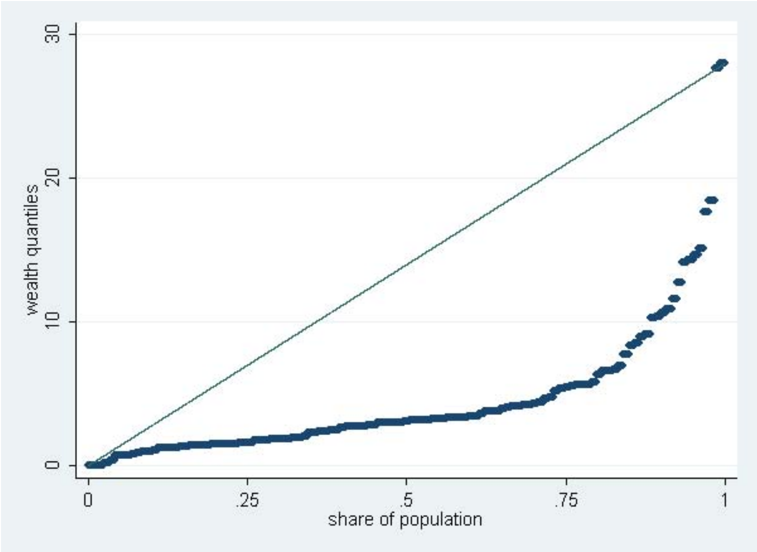

Table 1.A: Households' Attributes

religion	Muslim	24	n=119
	Lutheran	46	
	Catholic	49	
education	At least one member with primary school	93	n=119
	No member with primary school	26	
clan composition	1 household	11	n=119
	2 households	5	
	3 households	2	
	4 to 10 households	5	
	12 to 23 households	3	
at least one member engaged in	casual labor	57	n=116
	trade	45	
	crops	108	
	livestock	31	
	assets	8	
	processing	41	
	other off-farm	40	

Table 2.A: Distribution of Dichotomous Variables

Variable	Value	Dyads (n=14042)
link g_{ij}	1	980
	0	13062
kinship	1	218
	0	13824
same clan	1	1318
	0	12724
same religion	1	4974
	0	9068
neighbors	1	3462
	0	10580
same education	1	9074
	0	4968

Table 3.A: Distribution of Continuous Variables

variable	mean	min	max	s.d.
$wealth_j$	4.55	0	27.97	4.82
$two\ steps\ gains_{ij}$	16.91	-30.92	125.90	10.91
$three\ steps\ gains_{ij}$	-13.86	-76.66	97.06	13.23
$labour\ share\ cropping_i$	0.39	0	1	0.26
$labour\ share\ casual_i$	0.19	0	1	0.24
$labour\ share\ trade_i$	0.11	0	1	0.20
$difference\ wealth$	4.40	0	27.97	5.24

Note: values computed on the full sample of 14042 dyads.

Table 4.A: Marginal Effects

	(1)	(2)	(3)
$wealth_j$	0.0142	0.0050	0.0047
$two\ steps\ gains_{ij}$		-0.0017	-0.0014
$three\ steps\ gains_{ij}$			0.0003
$neighbours$	0.1231	0.0404	0.0354
$same\ clan$	0.0483	0.0152	0.0130
$Lutheran_i$	0.0152	-0.0043	-0.0030
$Catholic_i$	-0.0033	-0.0080	-0.0067
$same\ religion$	0.0376	0.0129	0.0113
$kinship$	0.3273	0.0844	0.0652
$same\ education$	-0.0018	0.0003	0.0006
$labour\ share\ cropping_i$	-0.0221	-0.0148	-0.0130
$labour\ share\ casual_i$	-0.0429	-0.0027	-0.0055
$labour\ share\ trade_i$	-0.0747	-0.0113	-0.0116
$difference\ wealth$	-0.0064	0.0003	0.0002
$more\ wealth_i$	0.5053	-0.0632	-0.0352

Table 5.A: Links by Wealth Quintile

quintile	mean wealth	average number of links
quintile 1	0.91	5.17
quintile 2	1.90	7.67
quintile 3	3.07	8.92
quintile 4	4.66	8.08
quintile 5	12.43	11.30
centile 90	16.82	12.55
centile 95	20.24	16.00

Note: 1 unit of wealth corresponds to 100000 *tzs*.